

TropGeneDB, the multi-tropical crop information system updated and extended

Chantal Hamelin*, Guilhem Sempere, Vincent Jouffe and Manuel Ruiz*

CIRAD, UMR AGAP, F-34398 Montpellier, France

Received August 14, 2012; Revised October 11, 2012; Accepted October 21, 2012

ABSTRACT

TropGeneDB (<http://tropgenedb.cirad.fr>) was created to store genetic, molecular and phenotypic data on tropical crop species. The most common data stored in TropGeneDB are molecular markers, quantitative trait loci, genetic and physical maps, genetic diversity, phenotypic diversity studies and information on genetic resources (geographic origin, parentage, collection). TropGeneDB is organized on a crop basis with currently nine public modules (banana, cocoa, coconut, coffee, cotton, oil palm, rice, rubber tree, sugarcane). Crop-specific Web consultation interfaces have been designed to allow quick consultations and personalized complex queries. TropGeneDB is a component of the South Green Bioinformatics Platform (<http://southgreen.cirad.fr/>).

INTRODUCTION

TropGeneDB is an information system initially developed by CIRAD (French Agricultural Research Centre for International Development) to manage various kinds of data on tropical crops (1).

TropGeneDB can record crop information on:

- molecular markers, quantitative trait loci (QTLs), genetic and physical maps
- genetic diversity studies
- phenotypic diversity studies based on agro-morphological traits or traits measuring susceptibility/resistance to aggression (diseases, salinity, drought)
- geographic origin, parentage, collection

The TropGeneDB project was initiated using the ACEDB database management system (2), but we migrated to MySQL to follow the evolution of computer technology. TropGeneDB is organized on a crop basis. Since the first version of TropGeneDB (1) containing three running modules (cocoa, banana and sugarcane), six additional modules have been implemented: coconut,

coffee, cotton, oil palm, rice and rubber tree. A sorghum module exists, but it is still in private access.

New crop-specific Web consultation interfaces have been designed to allow quick and complex queries with user-friendly result representation. Links to the GMOD CMAP viewer (the Comparative Map Viewer) (3) have been integrated.

All the data in TropGeneDB are public and generally linked to published scientific articles. Most of the data in TropGeneDB are original data not available in other Web information systems. Indeed, the TropGeneDB rice module stores unique genetic and phenotypic data on European collections of rice (4). The TropGeneDB banana module was identified by the Global Musa Genomics Consortium as a reference database for banana markers and genetic maps (5). No other equivalent database exists for most of our crop modules (coconut, cocoa, rubber tree, oil palm, sugarcane). Our resources contain data about important crops for the agrarian economy of many tropical countries. Indeed, these data can be exploited for rational use of genetic diversity available from germplasm collections, genome mapping and marker-assisted selection.

TropGeneDB is a component of the South Green Bioinformatics Platform (<http://southgreen.cirad.fr/>) and is accessible at <http://tropgenedb.cirad.fr>.

DATABASE CONTENTS

Currently, TropGeneDB contains ~19 800 molecular markers and 9500 germplasm entries (Table 1). Molecular data comprise genotypes at various types of markers (SNP, DArT, SSR, RFLP, AFLP, etc.), information on the markers themselves (probes, primers, sequences, etc.), QTLs, genetic and physical maps. Germplasm entries are linked to passport data (collections, accession identifier, country of origin, etc.) and much other detailed information: taxonomy, ploidy, ecosystem, etc. They are also linked to detailed phenotypic information based on agronomic and morphological traits and measurements of susceptibility or resistance to abiotic and biotic stress. Most of the rice trait data, used in

*To whom correspondence should be addressed. Tel: +33 4 67 61 59 75; Fax: +33 4 67 61 71 83; Email: chantal.hamelin@cirad.fr
Correspondence may also be addressed to Manuel Ruiz. Tel: +33 4 67 61 65 29; Fax: +33 4 67 61 56 05; Email: manuel.ruiz@cirad.fr

Table 1. Number of TropGeneDB entries (15 July 2012)

	Markers	Germplasms	QTLs	Map studies	Phenotyping studies
Banana	1068	541		5	
Cocoa	2108		187	8	
Coffee	258	2			
Coconut	471	174	63	1	
Cotton	5786	13	478	8	
Oil palm	969				
Rice	3466	2041	2179	1	26
Rubber tree	724	4		3	
Sugarcane	5026	6737	81	6	33
Total	19 876	9512	2988	32	59

phenotyping and QTL studies, are annotated with trait ontology terms (6).

WEB CONSULTATION INTERFACE

Query forms

Each crop module has several interfaces to carry out specific requests. The interfaces are tab-based, with each tab containing a different query form that leads to different types of data: phenotypes, genotypes, passport data, markers, QTLs, etc. A query form consists of various kinds of filters represented by standard input widgets. An important point is the ability to select the operator (AND/OR) applied for grouping together the filters defined by the used widgets. Filters corresponding to numerical values can be combined using relational operators (more than, equal to, etc.). For mapping data, users can filter on mapped markers and QTLs inside or overlapping a segment of given left and right limits (bp or cM) (Figure 1A).

Some text boxes have lookup lists tied to them. This is particularly useful when a non-trivial, exact string is expected. Filters where a red cross lies between the criterion label and the widget can be removed by clicking on the indicated cross icon. Those filters can also be (re-)added, several times if needed, using the 'Add criterion' list box. Having several identical filters only makes sense when using the OR operator because otherwise selecting two different values for the same criterion would result in an empty data set.

Filters for which no red cross appears primarily consist of list boxes whose contents are subject to a dependency relationship. This kind of relationship ties list boxes two by two: the contents of the second one are automatically updated when selection changes in the first. This hierarchical link can be between a genotyping study and the list of corresponding germplasms, or between a marker type and the corresponding list of markers. Furthermore, for rice phenotype and QTL studies, we have defined a standardized hierarchy of trait classes (Figure 1A).

Result data

Search results are displayed in tables where each column represents a relevant attribute (database field) (Figure 1B). Visualized data can be sorted by any of those fields by

clicking on the column header label. In some cases where additional details are available regarding a given attribute, its value is highlighted in orange: when moused-over, an extra layer appears where one or more links can be clicked. Indeed, the link 'Detail sheet' provides internal TropGeneDB details on the attribute (Figure 1C). External links are also available like Gramene (6) for the germplasms and markers of the rice TropGeneDB module. For maps, mapped markers and QTLs, links to a GMOD CMAP viewer (the Comparative Map Viewer) (3) have been integrated. The CMAP viewer provides a graphical representation of the correspondences between markers on different genetic and/or physical maps (Figure 1D). From the CMAP viewer, a link is enabled to the GMOD GBROWSE (the genome viewer) when right clicking a mapped object. This link is reciprocal. Pagination is used for displaying the results. Apart from being able to move to the first, next, previous or last page, users can also define the number of records per page, or directly jump to a page by specifying its index.

Exporting results

At the end of the navigation bar can be found an 'Export results' list box that lets you select an export format: Excel or Text (csv) (Figure 1E). Depending on the kind of data that are being dealt with, two extra formats may be offered: Excel matrix and Text matrix. Indeed, the results of queries on phenotype or genotype data, which respectively involve traits and germplasms, and markers and germplasms, are in row mode by default, i.e. one row for a combination of germplasms/traits or germplasms/markers. When exported in matrix format, those results are rebuilt as a matrix with the germplasms in rows and the traits or markers in columns.

DATA SUBMISSION

TropGeneDB data have been submitted by CIRAD teams and scientists from other institutions working on tropical crops: a data origin Web page is available for each crop module. People wishing to submit data may download submission templates (Microsoft Excel files), instructions for filling in the templates and can contact us using the Web form. Submitted data quality and integrity are checked by biologists. These biologists are researchers,

PHENOTYPES

GENOTYPES

PASSPORT_DATA

MARKERS

QTLs

Search now

Reset form

QTL study

Trait class

Trait

Environment

Water condition

Experiment location

Stress stage

Phenotypic R2 (%)

QTL inside a segment: segment left limit (bp)

QTL inside a segment: segment right limit

QTL overlapping a segment: segment left limit (bp)

QTL overlapping a segment: segment right limit (bp)

OPERATOR: AND OR

Add criterion

Search now

Reset form

Champoux et al (1995)

Courtois et al (2000)

Courtois et al (2003)

Dong et al (2004)

Gomez et al (2006)

DROUGHT: non root traits

DROUGHT: root traits

SALINITY: non root traits

SALINITY: root traits

ABA content

Biomass

Break point of osmotic adjustment

Canopy temperature

Anaerobic

not like

like

like

like

>

<

>

<

(A) Query filters

RESULTS : 4 records found

<< < Page n° 1 / 1 > >>

Number of records per page 50

[Export results]

QTL study	Trait	Map	Chromosome	QTL	Left locus	Right locus	Conditions	Phenotypic R2 (%)	Sampling method	QTL start position (bp)	QTL stop position (bp)	Population name	Population type	Population size
Courtois et al (2003)	Biomass	Detailed sheet	9	RG553	RZ206		Greenhouse; PVC pipes; aerobic conditions; well watered treatment	5.50	1 plant; 45 das	9216376.00	11807649.00	IAC165/Co39	RILs F7	125
Courtois et al (2003)	Biomass	SYNTHETIC_MAP1 5	9	RM39	RM233B		Greenhouse; PVC pipes; aerobic conditions; well watered treatment	6.20	1 plant; 45 das	20672421.00	23795831.00	IAC165/Co39	RILs F7	125
Courtois et al (2003)	Biomass	SYNTHETIC_MAP1 4	9	RG214	RG620		Greenhouse; PVC pipes; aerobic conditions; well watered treatment	24.80	1 plant; 45 das	31662839.00	34698383.00	IAC165/Co39	RILs F7	125
Courtois et al (2003)	Biomass	SYNTHETIC_MAP1 1	9	RZ649B	RZ801		Greenhouse; PVC pipes; aerobic conditions; well watered treatment	12.10	1 plant; 45 das	39433049.00	40566030.00	IAC165/Co39	RILs F7	125

(B) Query results

Reference Rice Synthetic_Map1

9

ES553

TEL95

RG553

RG214

RG620

RZ649B

RZ801

TEL96

ES554

ES555

ES556

ES557

ES558

ES559

ES560

ES561

ES562

ES563

ES564

ES565

ES566

ES567

ES568

ES569

ES570

ES571

ES572

ES573

ES574

ES575

ES576

ES577

ES578

ES579

ES580

ES581

ES582

ES583

ES584

ES585

ES586

ES587

ES588

ES589

ES590

ES591

ES592

ES593

ES594

ES595

ES596

ES597

ES598

ES599

ES600

ES601

ES602

ES603

ES604

ES605

ES606

ES607

ES608

ES609

ES610

ES611

ES612

ES613

ES614

ES615

ES616

ES617

ES618

ES619

ES620

ES621

ES622

ES623

ES624

ES625

ES626

ES627

ES628

ES629

ES630

ES631

ES632

ES633

ES634

ES635

ES636

ES637

ES638

ES639

ES640

ES641

ES642

ES643

ES644

ES645

ES646

ES647

ES648

ES649

ES650

ES651

ES652

ES653

ES654

ES655

ES656

ES657

ES658

ES659

ES660

ES661

ES662

ES663

ES664

ES665

ES666

ES667

ES668

ES669

ES670

ES671

ES672

ES673

ES674

ES675

ES676

ES677

ES678

ES679

ES680

ES681

ES682

ES683

ES684

ES685

ES686

ES687

ES688

ES689

ES690

ES691

ES692

ES693

ES694

ES695

ES696

ES697

ES698

ES699

ES700

ES701

ES702

ES703

ES704

ES705

ES706

ES707

ES708

ES709

ES710

ES711

ES712

ES713

ES714

ES715

ES716

ES717

ES718

ES719

ES720

ES721

ES722

ES723

ES724

ES725

ES726

ES727

ES728

ES729

ES730

ES731

ES732

ES733

ES734

ES735

ES736

ES737

ES738

ES739

ES740

ES741

ES742

ES743

ES744

ES745

ES746

ES747

ES748

ES749

ES750

ES751

ES752

ES753

ES754

ES755

ES756

ES757

ES758

ES759

ES760

ES761

ES762

ES763

ES764

ES765

ES766

ES767

ES768

ES769

ES770

ES771

ES772

ES773

ES774

ES775

ES776

ES777

ES778

ES779

ES780

ES781

ES782

ES783

ES784

ES785

ES786

ES787

ES788

ES789

ES790

ES791

ES792

ES793

ES794

ES795

ES796

ES797

ES798

ES799

ES800

ES801

ES802

ES803

ES804

ES805

ES806

ES807

ES808

ES809

ES810

ES811

ES812

ES813

ES814

ES815

ES816

ES817

ES818

ES819

ES820

ES821

ES822

ES823

ES824

ES825

ES826

ES827

ES828

ES829

ES830

ES831

ES832

ES833

ES834

ES835

ES836

ES837

ES838

ES839

ES840

ES841

ES842

ES843

ES844

ES845

ES846

ES847

ES848

ES849

ES850

ES851

ES852

ES853

ES854

ES855

ES856

ES857

ES858

ES859

ES860

ES861

ES862

ES863

ES864

ES865

ES866

ES867

ES868

ES869

ES870

ES871

ES872

ES873

ES874

ES875

ES876

ES877

ES878

ES879

ES880

ES881

ES882

ES883

ES884

ES885

ES886

ES887

ES888

ES889

ES890

ES891

ES892

ES893

ES894

ES895

ES896

ES897

ES898

ES899

ES900

ES901

ES902

ES903

ES904

ES905

ES906

ES907

ES908

ES909

ES910

ES911

ES912

ES913

ES914

ES915

ES916

ES917

ES918

ES919

ES920

ES921

ES922

ES923

ES924

ES925

ES926

ES927

ES928

ES929

ES930

ES931

ES932

ES933

ES934

ES935

ES936

ES937

ES938

ES939

ES940

ES941

ES942

ES943

ES944

ES945

ES946

ES947

ES948

ES949

ES950

ES951

ES952

ES953

ES954

ES955

ES956

ES957

ES958

ES959

ES960

ES961

ES962

ES963

ES964

ES965

ES966

ES967

ES968

ES969

ES970

ES971

ES972

ES973

ES974

ES975

ES976

ES977

ES978

ES979

ES980

ES981

ES982

ES983

ES984

ES985

ES986

ES987

ES988

ES989

ES990

ES991

ES992

ES993

ES994

ES995

ES996

ES997

ES998

ES999

ES1000

Trait

Description

TO term

TO ID

Trait acronym

Biomass

Sum of plant shoot dry weight and grain dry weight

Biomass

TO:0000327

BIOM

(C) Detail sheet

TropGene DB			QTLs								
09/08/2012			17:29:01								
QTL study			Trait	Map	Chromosome	QTL	Left locus	Right locus	Conditions	Phenotypic R2 (%)	
6	Courtois et al (2003)	Biomass	SYNTHETIC_MAP1	9	148	RG553	RZ206		Greenhouse; PVC pipes; aerobic conditions; w 5.50		
7	Courtois et al (2003)	Biomass	SYNTHETIC_MAP1	5	147	RM39	RM233B		Greenhouse; PVC pipes; aerobic conditions; w 6.20		
8	Courtois et al (2003)	Biomass	SYNTHETIC_MAP1	4	146	RG214	RG620		Greenhouse; PVC pipes; aerobic conditions; w 24.80		
9	Courtois et al (2003)	Biomass	SYNTHETIC_MAP1	1	145	RZ649B	RZ801		Greenhouse; PVC pipes; aerobic conditions; w 12.10		

(E) Data export

(D) CMAP view

Figure 1. Example of a rice QTL query and corresponding results. (A) Query filters, (B) Query results, (C) Detail sheet, (D) CMAP view, (E) Data export.

experts for a given crop, who help the database administrator in checking crucial points: for instance, in standardizing the names of the germplasms, the markers and the traits. They also assess the relevance of the submitted studies.

DESIGN AND IMPLEMENTATION

TropGeneDB is a Web application based on MySQL databases, with one per crop that can be queried using Java customizable interfaces automatically generated to fit the database contents. A generic database model has been developed and is used for all the crop modules.

CONCLUSION

The TropGeneDB database is a long-term project. This database is permanently updated: the range of crops will soon be broaden to alfalfa, einkorn, olive tree, pearl millet, sorghum, tomato. New data types, such as association results and linkage disequilibrium, will also be added.

The user interface could be improved: the input widgets could be placed differently to prevent the need of scrolling up and down to see them all when they are numerous.

AVAILABILITY

Authors who use and download data from TropGeneDB are encouraged to cite this article.

ACKNOWLEDGEMENTS

The authors should like to thank all the expert biologists involved in TropGeneDB data curation. They also thank Gaëtan Droc for the reciprocal links between CMAP and GBrowse. They are grateful to Peter Biggins for help in

English editing. C.H. and V.J. implemented the new version of the MySQL database, developed the data submission templates and the Perl scripts for data insertion. G.S. developed the Java Web interfaces. C.H. managed data updates and the design of the Web query forms. M.R. coordinated the project and wrote the article.

FUNDING

CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement); International Consortium for Sugarcane Biotechnology (to V.J.); Agropolis Foundation through the Agropolis Resource Centre for Crop Conservation, Adaptation and Diversity project (ARCAD) (to G.S.). Funding for open access charge: CIRAD.

Conflict of interest statement. None declared.

REFERENCES

1. Ruiz,M., Rouard,M., Raboin,L.M., Lartaud,M., Lagoda,P. and Courtois,B. (2004) TropGENE-DB, a multi-tropical crop information system. *Nucleic Acids Res.*, **32**, D364–D367.
2. Kelley,S. (2000) Getting started with Acedb. *Brief. Bioinform.*, **1**, 131–137.
3. Youens-Clark,K., Faga,B., Yap,I.V., Stein,L. and Ware,D. (2009) CMap 1.01: a comparative mapping application for the Internet. *Bioinformatics*, **25**, 3040–3042.
4. Courtois,B., Frouin,J., Greco,R., Bruschi,G., Droc,G., Hamelin,C., Ruiz,M., Clément,G., Evrard,J.C., van Coppenole,S. *et al.* (2012) Genetic diversity and population structure in a european collection of rice. *Crop Sci.*, **52**, 1663–1675.
5. Rouard,M., Carpentier,S.C., Bocs,S., Droc,G., Argout,X., Roux,N. and Ruiz,M. (2012) In: Pillay Michael,U.G.K.C. (ed.). Science Publishers, Enfield, pp. 194–216.
6. Youens-Clark,K., Buckler,E., Casstevens,T., Chen,C., Declerck,G., Derwent,P., Dharmawardhana,P., Jaiswal,P., Kersey,P., Karthikeyan,A.S. *et al.* (2011) Gramene database in 2010: updates and extensions. *Nucleic Acids Res.*, **39**, D1085–D1094.